≡ Sections

Option C - Web Science (/topic/C-web-science.html)

C.1 Creating the web

Evolution of the web, different protocols and web technologies. Difference between static and dynamic web pages. External data sources. Role of the browser.

C.1.1 Distinguish between the Internet and World Wide Web

Internet

- interconnected set of networks and computers
- Permits transfer of data(e.g.: WWW, email, P2P, VOIP, FTP)
- Permits delivery of services
- Data transfer governed by protocols(TCP/IP)

World Wide Web

- A set of hypertext-linked resources
- Resources are identified by URIs(Unified Resource Identifier)
- Transferred between client and server via the Internet
- Resources can be read by use of a browser

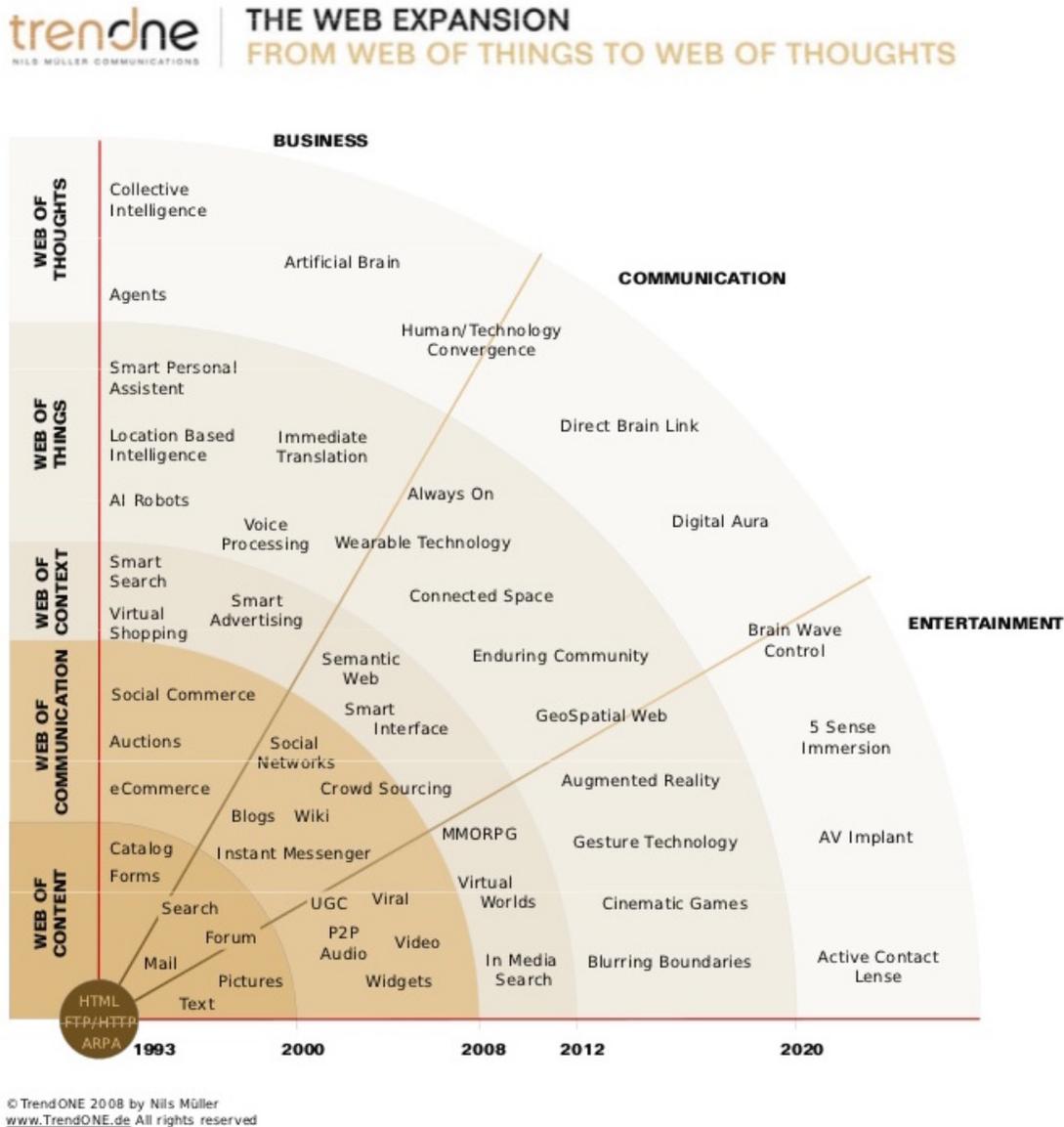## C.1.2 Describe how the web is constantly evolving

Figure 1: The Web Expansion (TrendOne, 2008 (http://www.slideshare.net/trendone/the-web-expansion))

The beginnings of the web (Web 1.0 , Web of content)

The world wide web started around 1990/91 as a system of servers connected over the internet that deliver static documents, which are formatted as hypertext markup language (HTML) files, which support links to other documents, but also multimedia as graphics, video or audio. In the beginnings of the web, these documents consisted mainly of static information and text, where multimedia were added later. Some experts describe this as a "read-only web", because users mostly searched and read information, while there was little user interaction or content contribution.

Web 2.0 – "Web of the Users"

However, the web started to evolve into the delivery of more dynamic documents, enabling user interaction or even allowing content contribution. The appearance of blogging platforms as Blogger in 1999 gives a time mark for the birth of the Web 2.0. Continuing the model from before, this would be the evolution to a "read-write" web. This opened new possibilities and lead to new concept as blogs, social networks or video-streaming platforms. Web 2.0 might also be looked at from the perspective of the websites themselves evolving in more dynamic and feature-rich. For instance, improved design, JavaScript and dynamic content loading could be considered Web 2.0 features.

Web 3.0 – "Semantic Web"

The internet and thus the world wide web is constantly developing and evolving into new directions and while the changes described for the Web 2.0 are clear to us today, the definition for the Web 3.0 is not definitive yet. Continuing the read to read-write description form earlier, it might be argued that the Web 3.0 would be the "read-write-execute" web. One interpretation of this, is that the web enables software agents to work with documents by using semantic markup. This allows for smarter searches and the presentation of relevant data fitting into context. This is why Web 3.0 is sometimes called the semantic executive web.

But what does this mean?

It's about user input becoming more meaningful, more semantic, by users giving tags or other kinds of data to their document, that allow software agents to work with the input, e.g. to make it more searchable. The idea is to be able to better connect information that is semantically connected.

Later developments

However, it might also be argued that the Web 3.0 is what some people call the Internet of Things, which is basically connecting every day devices to the internet to make them smarter. In some way, this also fits the read-write-execute model, as it allows the user to control a real life action on a device over the internet. Either way, the web keeps evolving and the following image provides a good overview and an idea where the web is heading to.

C.1.3-4 Identify the characteristics of the following:

HTTP – Hypertext Transfer Protocol

- Application layer protocol from the Internet Protocol suite to transfer and exchange hypermedia
- request-response protocol based on client-server model
- user agent (e.g. web browser) requests some resource from a server through an URL, and the web server gives and response

- different HTTP request methods, e.g. for retrieving or submitting data (GET and POST)

HTTPS – Hypertext Transfer Protocol Secure

- Based on HTTP
- Adds an additional security layer of SSL or TLS
- ensures authentication of website by using digital certificates
- ensures integrity and confidentiality through encryption of communication
- still possible to track IP address and port number of web server (which is why HTTPS websites are also blocked in China)

HTML – Hypertext Mark-up Language

- semantic markup language
- standard language for web documents
- uses elements enclosed by tags to markup a document

URL – Uniform Resource Locator

- unique string that identifies a web resource
- reference to a web resource
- primarily used for HTTP, but also for other protocols like FTP or email (mailto)
- includes name AND access method (e.g. 'http://')
- serves as a mechanism to retrieve a resource
- follows a specific syntax
  schema: `[//[user:password@]host[:port]][/]path[?query][#fragment]`

XML – Extensible Mark-up Language

- markup language with a set of rules defining how to encode a document
- human-readable
- similar to HTML in using tags
- used for representation of arbitrary data structures

XLST – Extensible stylesheet language

- styling language for XML used for data presentation and transformation
- data presentation means displaying data in some format/medium, about style
- data transformation is about parsing a source tree of nodes out and transform it into something different
- XLST can be used to transform XML files into other XML files, HTML, PDF, PNG and others

JavaScript

- interpreted programming language
- core technology of most websites with HTML and CSS

- high-level, dynamic and untyped; therefore relatively easy for beginners
- allows to dynamically manipulate the content of HTML documents
- makes websites dynamic

CSS – Cascading style sheet

- style sheet language to describe the presentation of a mark-up document, usually HTML
- used to create better designed websites
- intended to separate content in presentation in HTML and CSS
- it uses selectors to describe particular elements of a document, and gives these properties that define things ranging from font color to page position

URI – Uniform Resource Identifier

- more general definition than URL
- a string serving as an identifier for some resource(document, image, mailbox, video, files, etc.)
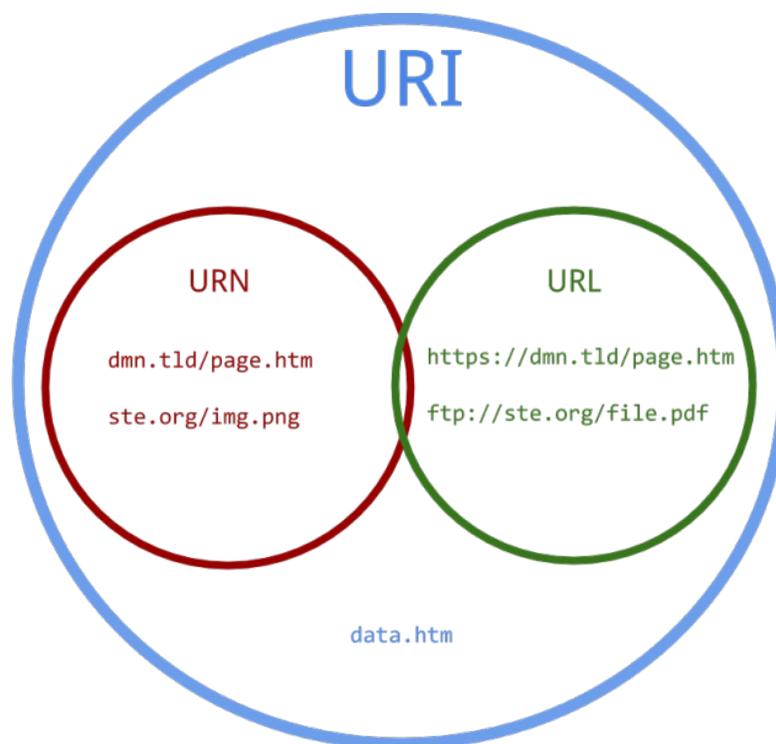

Figure 2: The Difference Between URLs and URIs (Daniel Miessler (https://danielmiessler.com/study/url-uri))

C.1.5 Describe the purpose of a URL

An URL defines a pathway to a specific resource. It allows to link different resources, which creates the basis for navigating the WWW (it is the links that make it a web).

C.1.6 Describe how a domain name server functions

A DNS server makes sure that a website can be accessed under a human-readable name like `www.cs-ib.net` . In order to access a website, such a name first needs to be translated into a machine-readable IP address like 52.219.32.34. The DNS server does exactly this. It stores the respective IP address corresponding to a domain name in its databases and returns it on request.

Further Reading: How Stuff Works - DNS Server (http://computer.howstuffworks.com/dns.htm)

C.1.7 Identify the characteristics of:

IP - (Internet Protocol):

- Part of the Internet layer TCP/IP Internet protocol suite
- Responsible for delivering packets to the correct address/computer
- Defines the format of a packet
- Adds a header to each TCP segment which includes routing information
- IPv4 uses a 32-bit address
- IPv6 will use 128-bit address

TCP – (Transmission Control Protocol):

- Part of the trabsport layer TCP/IP Internet protocol suite
- Receives data from an applicaiton and divides it up (into segments) ready for IP
- Establishes initial connection
- Provides error-checking which can lead to the re-sending of packets

FTP - (File Transfer Protocol):

- Protocol for transferring files (between server and client) over a TCP based network
- Files can also be manipulated/modified remotely
- Control information(e.g.: log-ins) are sent separately from the main file data(out-of-band control) which differs FTP from HTTP

C.1.8 Outline the different components of a web page

A web page can contain a variety of components. The basics structure of a HTML document is:

```
head
```

This is not visible on the page itself, but contains important information about it in form of metadata.

`title`

The title goes inside the head and is usually displayed in the window top of the web browser.

`meta tags`

There are various types of meta tags, which can give search engines information about the page, but are also used for other purposes, such as to specify the charset used.

`body`

The main part of the page document. This is where all the (visible) content goes in.

Some other typical components:

Navigation bar

Usually a collection of links that helps to navigate the website.

Hyperlinks

A hyperlink is a reference to another web page.

Table Of Contents

Might be contained in a sidebar and is used for navigation and orientation within the website.

Banner

Area at the top of a web page linking to other big topic areas.

Sidebar

Usually used for a table of contents or navigation bar.

Corner

Most of the time used for logo with a link to the main page.

Continuation

Area of the web page preventing the sidebar to extend to the bottom of the web page.

C.1.9 Explain the importance of protocols and standards on the web

Protocols

- A set of rules for successfully carrying out some process
- Investigate TCP/IP/FTP

Standards

- Set of technical specifications that should be adhered to, to allow for functionality
- Web page / browser standards include correct use of HTML, JavaScript, XML, CSS etc.
- Semantic web standards involve the use of ontologies
- Interoperability
- Accessibility

What is ISO?

ISO is the International Organization of Standardization, an independent, non-governmental organization that develops and publishes international standards. It has been founded in 1946 and since then has published over 21000 international standards regarding aspects of technology and manufacturing.

Who constitutes ISO?

The members are from 163 countries including 3 368 technical bodies that help standards to be developed. In addition, the organization has over 135 people working fulltime at the central in Geneva. Experts of the same field work together to develop standards and these are settled on through a consensus process.

Why is there a need for ISO?

These standards ensure safety, reliability and quality for products and services, while also providing a common denominator for different processes to communicate, e.g. for technologies.

C.1.10 Describe the different types of web page

Personal Page

Characteristics:

- Site about individuals/small groups
- Includes information the creator wants to

Technical Aspects:

- Mainly displays content, therefore usually static
- Usually does not require any back-end server-side programming (static)
- May use some front end technologies
- Often not self-hosted and not coded, but created using a WYSIWYG website creator like wix, Dream Weaver

Blog

Characteristics:

- Site about certain topics made by an individual
- The "blogger" publishes smaller contents most commonly shown the most recent updated posts
- Is a sort of online diary

Technical Aspects:

For a static blog:

- HTML
- CSS
- Optionally, static content manager, e.g. Jekyll

For a dynamic blog:

- Database, e.g. MySQL, MongoDB
- JavaScript for asynchronous communication with database, e.g. loading comments
- Backend programming language, e.g. PHP, to handle incoming posts, etc.
- Content Management System (CMS), e.g. WordPress

Search Engine Pages

Characteristics:

- Indexes content from the internet or an intranet
- Serves related links based on a users query

Technical Aspects:

- Use a web crawler to maintain real-time information –> looks at meta tags of websites
- Back-end is programmed in an efficient language, e.g. C++ HTML, CSS, JavaScript

Forums

Characteristics:

- A meeting or medium where ideas and views on a particular issue can be exchanged

Technical Aspects:

- Database for storing posts
- Some language, e.g. PHP, Node.js, Ruby on Rails, Django
- CMS (content management system), e.g. vBulletin
- classic client side, HTML, CSS, JavaScript

Wiki

Characteristics:

- A website or database developed by many different users
- Any user can add and edit content

Technical Aspects:

- HTML
- CSS
- JavaScript
- Dynamic (accessing Database)
- content management system (CMS), e.g. mediaWiki

News Site

Characteristics:

- A website that offers current news in form of articles or videos
- Includes updated news feed

Technical Aspects:

- usually dynamic
- Content management system (CMS)
- HTML, CSS, JavaScript
- Usually some database
- back end programming language

Photo-sharing site

Characteristics:

- Site that allows free photo sharing by online advertising

Technical Aspects:

- Examples: Flickr.com, Picasa
- Software that enables digital photo slide shows and then upload these to the web
- HTML, CSS, JavaScript

Learning Management System

Characteristics:

- Software application for delivery, administration, tracking, documentation, reporting of e-learning courses.

Technical Aspects:

- Heavy client-side programming for interactive sites (JavaScript) for exercise checking, file upload, videos
- Back-end programming + database to store students progress, course contents, etc.

C.1.11 Explain the differences between a static web page and a dynamic web page

Static Websites

Sites that only rely on the client-side and don't have any server-side programming. This means that all content is available through the HTML, CSS and JavaScript files served to the client. This way the server doesn't do anything else than serving files. The website can still be dynamic in the sense of JavaScript manipulating the available content, e.g. for animations and such.

Advantages:

- Lower cost to implement
- flexibility

Disadvantages:

- Low scalability
- Hard to update
- Higher cost in the long term if content is to be updated

Dynamic Websites

Sites that include server-side programming as well, usually to retrieve content dynamically from a database. This allows for data processing on the server and allows for much more complex applications.

Advantages:

- information can be retrieved in an organized way from databases
- allows for content management systems
- low ongoing cost, unless design changes or extra features are implemented

Disadvantages:

- sites are usually based on templates, which information is fed into
  - less individual sites
- higher initial cost
- usually larger codebase

C.1.12 Explain the functions of a browser

- Interprets and displays information sent over the Internet in different formats(e.g. HTTP/applications/AJAX/JavaScript)
- Plug-ins allows multimedia to be displayed
- Retrieves information from the Internet via hyperlinks

C.1.13 Evaluate the use of client-side scripting and server-side scripting in web pages

Server-side

Website logic that runs on the server. Common tasks include the processing of search queries, data retrieval from a database and various data manipulation tasks. A good example are online-shops, where items are displayed based on a search query. Once the user decides to buy an item, server-side scripts check user credentials and make sure that the shop receives the order.

Technologies include:

- CGI
- Direct execution(e.g. ASP, PHP)

Client-side

This is scripting that happens in the browser of the client. It is used for animations, form validation and also to retrieve new data without reloading the page, e.g. in a live-chat.

Some technologies:

- JavaScript
- AJAX
- JQuery

C.1.14 Describe how web pages can be connected to underlying data sources

Client-side

Cookies: Cookies are small files stored on a user's computer. They hold data specific to a website or client and can be accessed by either the web server or the client computer. Cookies contain data values such as first-name and last-name. Once the server or client computer have read the cookie through their respective codes, the data in the cookie can be retrieved and used for a website page. Cookies are created usually when a new web page is loaded. Disabling cookies on your computer will abort the writing operation that creates cookies. However, some sites require cookies in order to function. Cookies are used to transport information from one session on a website to another. They eliminate the use of server machines with huge amounts of data storage, since cookies are more efficient and smaller.

Server-side

Databases: A database is an organized collection of data, which allows to retrieve specific data easily based on queries. Data are usually organized in a way that allows the application to find data easily. There is different logical models of how to organize data in a database, e.g. relational models, object models, navigational models and more. A database is accessed (in order to retrieve data, update them, administration) through a database management system (DBMS), such as for example MySQL, PostgreSQL, MongoDB, etc. . These systems usually differ in the database model that they use. The DBMS usually provides some sort of library through which scripts in various languages (e.g. PHP, JavaScript, ASP.NET) can make queries and read or manipulate data.

XML: XML is a flexible way to structure data and can therefore be used to store data in files or to transport data. It allows data to be easily manipulates, exported, or imported. This way, websites can also be designed independent from the data content. An example use of XML are RSS feeds, where it's used to store data about a feed.

CGI(Common Gateway Interface): See C.1.15

C.1.15 Describe the function of the common gateway interface (CGI)

This is a standard protocol for web servers to execute console programs (applications that run from the command line) in order to generate dynamic websites. It implements an interface for the web server (as in the software) to pass on user information, e.g. a query, to the application, which can then process it. This passing of information between the web server and the console application is called the CGI. Thanks to CGI, a variety of programming languages such as Perl, Java, C or C++ can be used, which allow for fast server-side scripting.

C.1.16 Evaluate the structure of different types of web pages

See evaluation from the previous sections, specially C.1.8, C.1.10 and C.1.11 .

Sources:

- Flat World Business - Evolution of the Web (https://flatworldbusiness.wordpress.com/flat-education/previously/web-1-0-vs-web-2-0-vs-web-3-0-a-bird-eye-on-the-definition/)
- Wikipedia - 'Website' (https://en.wikipedia.org/wiki/Website)
- Website Types - Xislegraphix.com (http://www.xislegraphix.com/website-types.html)
- Blog Basics - Quora (https://www.quora.com/I-want-to-make-a-blog-I-know-the-basics-of-HTML-CSS-and-JavaScript-to-build-my-blog-How-should-I-start-writing-the-codes-or-what-will-be-the-steps)
- Difference Static/Dynamic Websites - Quora (https://www.quora.com/What-is-the-difference-between-Static-Websites-and-Dynamic-Websites)
- Static/Dynamic Websites - Spiderwriting.co.uk (http://www.spiderwriting.co.uk/static-dynamic.php)
- What are cookies? (http://www.whatarecookies.com)
- CGI (http://searchsoa.techtarget.com/definition/common-gateway-interface)

Home (https://www.cs-ib.net)   About (/about)   Privacy policy (/privacy)